IP Quality of Service

Markus Peuhkuri
Helsinki University of Technology, Laboratory of Telecommunications Technology

May 10, 1999

Abstract

The quality of service is one of most important areas of Internet development. As the Internet originally developed for data communications is now used more and more for real-time applications, there is a need for better service than best effort. In this study we will first review the concept *quality of service*, what it is and then we study the two most important efforts to provide QoS in the Internet: the Integrated Services and the Differentiated Services models. Both have their own application area as the first provides more fine grained control to network resources as the later gives better scalability.

Contents

Other author information: Email: *Markus.Peuhkuri@hut.fi*; Telephone: +358-9-451 2467; Fax: +358-9-451 2474; Home page: URL:*http://www.iki.fi/puhuri/*.

1 Introduction

The Internet has recently became important communications channel. As it was used in 1980s and beginning of 1990s by research and education communications for computer data transmission: electronic mail, network news (Usenet) and file transfers. The most demanding application from service quality was network virtual terminal (telnet [44]) as it was an interactive application. The bandwidth required was small and occasional delay variations of order of several seconds could be tolerated [36].

Recently many interactive or real-time services has been introduced and at same time the economical importance of the Internet has grown. The IP phones and services based on that technology is threatening traditional circuit-switched telephone services, specially on long-distance applications.

Transmitting interactive real-time media is the greatest challenge in packet based networks. The end-to-end delay, the delay variations (jitter), and the packet loss must not exceed some limits or usability of the service degrades badly [32].

1.1 Why to need any QoS?

The Internet has worked so far with *best effort* traffic model: every packet is treated (forwarded or discarded) equally. This is very simple and efficient model and several arguments has been stated against any need for more complicated system [13]:

"Bandwidth will be infinite." The optical fiber has enormous transmission capacity, tens if not hundreds of terabits per second in a single 0.5 mm thin (including primary coating) strain of fiber. However, installing new capacity and developing faster equipment will take some time. Also, networks are generally designed in cost effective manner balancing between over-engineering and over-subscribing.

As wireless systems are more and more common, they limit available bandwidth because there are no much extra capacity on usable radio frequencies. Also energy conservation in portable equipment may limit available bandwidth.

> *Corollary of Moore's Law:* As you increase the capacity of any system to accommodate user demand, user demand will increase to consume system capacity. [21, p. 10]

"Simple priority is sufficient." This is very much true: QoS is all about giving some traffic higher priority over other traffic. The problem is where to assign the priority. User terminals cannot generally be trusted to give "fair" priorities for different services. If there is some billing and policing mechanism, then we do have already some kind of QoS mechanism. In some cases it is useful to give user busy signal to protect the network from being over-subscribed.

There are two approaches to assign priority in Internet traffic: hop-by-hop based on reservation (Integrated services, Chapter 4) and packet marking at edges of network (Differentiated services, Chapter 5).

"Applications can adapt." While the applications and protocols can adapt to even extreme delays, user adaptation is much less. For example, to maintain dialogue in telephone conversation, end-to-end delays cannot exceed 300 ms to avoid man-on-moon effect [32].

2 Quality of Service

The Quality of Service (QoS) is often quite much abused term. If we look at traditional circuit switched telecommunication networks, the QoS is formed by several factors, which can be divided into two groups: "human" and "technical" factors as shown in Table 1.

2.1 QoS in Packet Switched Networks

In packet switched network there are much more factors that must be agreed on. The Asynchronous Transfer Mode (ATM) networks have very extensive QoS control as it is intended for real-time traffic [38]. For IP networks the ITU is developing recommendation I.380 [25] which defines quite similar metrics for IP packet transfer performance parameters:

Table 1: Quality of service factors [45].

Human factors	Technical factors
stability of service quality	reliability
availability of subscriber lines	expandability
waiting times	effectiveness
fault clearance times	maintainability of the system
subscriber information	congestion waiting
stability of operation of the system	transmission quality
...	...

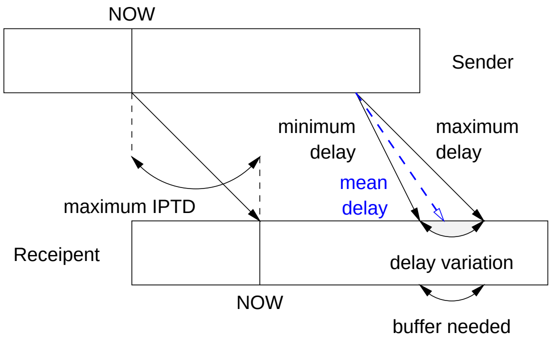

Figure 1: Difference of “current time” between receiver and sender.

IP packet transfer delay (IPTD) This is the delay for IP datagram (or delay for last fragment) between two reference points. Typically end-to-end delay or delay within one network.

Mean IP packet transfer delay An arithmetic average of IP packet transfers delays for packets we are interested about.

IP packet delay variation It is useful that streaming applications know how much the delay varies in network to avoid buffer overflows and underflows (Figure 1). For elastic applications (like TCP, see Chapter 3.1) small delay variations are not important but large ones may cause either unnecessary packet retransmissions or unnecessary long delays before retransmit.

IP packet error ratio (IPER) This is a ratio of errored packets of all received packets.

$$IPER = \frac{N_{\text{erroneous}}}{N_{\text{successful}} + N_{\text{erroneous}}} \tag{1}$$

IP packet loss ratio (IPLR) The ratio of lost packets from all packets transmitted in population of interest.

$$IPLR = \frac{N_{\text{lost}}}{N_{\text{transmitted}}} \tag{2}$$

The packet loss ratio affects on quality of connection. How the application reacts on packet loss can be divided in three categories (Figure 2). The applications can be divided to similar categories also by how much they require bandwidth.

Fragile If the packet loss exceeds certain threshold, the value of application is lost.

Tolerant The application can tolerate packet loss, but the higher the packet loss the lower is the value of application. There are certain threshold levels which are critical.

Performance The application can tolerate even very high packet loss ratio but its performance can be very low in high packet loss ratio.

Spurious IP packet rate As it is not expected that is number is proportional to number of packets transmitted, this is expressed as rate: number of spurious packets in time interval.

For ATM networks there are also metrics to characterize traffic flow for call admission control (CAC) purposes:

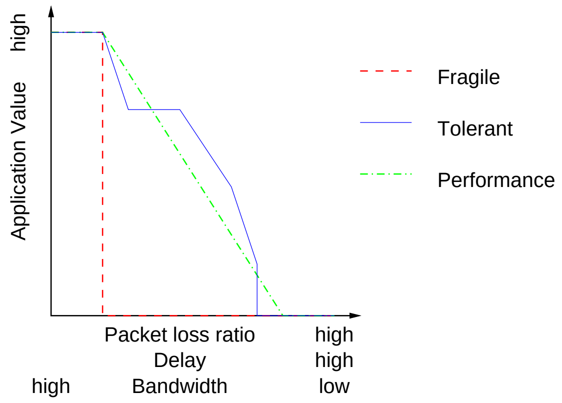

Figure 2: Application types.

Table 2: Differences between IPPM and ITU-T regarding time metrics [37]

IPPM	ITU-T	Definition
synchronization	time error	difference of two clocks
accuracy	time error from UTC	difference to real time
resolution	sampling period	the precession of clock
skew	time drift	change in synchronization or in accuracy

Peak Cell Rate (PCR) The maximum cell rate that connection may have while maintaining jitter less than defined by Cell Delay Variation Tolerance (CDVT).

Sustainable Cell Rate (SCR) The long-term maximum cell rate the connection may have.

Maximum Burst Size (MBS) The number of cells in burst which may exceed SCR but not PCR.

Minimum Cell Rate (MCR) The minimum rate the connection must be able to send at any time.

2.2 Efforts to define QoS

The IETF Internet Protocol Performance Metrics (IPPM) working group is working to define metrics for Internet performance. The framework document [37] defines criteria for those metrics, terminology, the metrics itself, methodology, and practical considerations including sources of uncertainty and errors. There are some differences in terminology considering time between ITU-T definitions and IPPM working group definitions. Short summary of differences is presented in Table 2.

As full traffic analysis is not always feasible, the IPPM metrics are based on random sampling of traffic. The framework document [37] includes discussion that recommends that Internet properties are not considered in probabilistic terms as there is no static state in Internet.

2.3 Grade of Service

The Grade of Service (GoS) has been used in telecommunications industry to indicate components which contribute to overall quality of service what the user receives. Many components have both human component and technical component: the technical component can be measured (like bandwidth of voice) and the human component is subjective. There is relation between human and technical component but the exact mapping depends on many factors, for example language used and other culture factors [45]. In [23] the GoS is defined as following:

> It may happen that in a network, or in part of a network, the volume of telephone traffic that arises exceeds the capacity for handling it without limitations, with the result that congestion occurs. These limitations affect the service provided to customers, and the degree of these limitations is expressed by an appropriate GoS parameter (e.g. probability of loss, average delay, failure rate due congestion, etc.). GoS should therefore be regarded as providing information on the traffic aspects of the "quality of service".

In circuit-switched network the GoS has been divided into two standards [28]:

Loss grade of service This standard has component *internal loss probability*: for any call attempt, it is the probability that an overall connection cannot be set up between a given incoming circuit and any suitable free outgoing circuit within the switching network. For international digital telephone exchanges the internal loss probability may not exceed 0.2 % in normal load situation and 1 % in high load. The figures are higher for end-to-end connections: mean 2 and 5 % in local and international connections respectively.

Delay grade of service There are several components in this standard, depending on technology used for signaling information. For the ISDN circuit-switched services the delay components are defined for pre-selection, post-selection, answer signal, and call release [29].

2.4 Class of Service

The Class of Service concept divides network traffic into different classes and provides class-dependent service to each packet depending on what class it belongs to. While the strict QoS has some absolute measures for quality, the CoS has relative measures: at this time this class gets packet drop probability of 10^{-6} while on the other class packet drop probability is 10^{-3}.

To differentiate the network traffic into different classes, the differentiation must be based on some factor. The factors include[21, p. 21]:

Protocol Differentiation is made based on some of protocols. The protocol information may be more or less accurate and includes:

- **Protocol identifier** One can differentiate IP from other network level protocols using link level information, TCP from UDP and ICMP using protocol field on IP header.
- **Source port number** The only way to identify applications[1] run over TCP or UDP is to look for port numbers and compare them to list of well-known port numbers, maintained by IANA/ICANN. While in most cases the mapping is correct there are many cases when some service or client uses port reserved for another application.

  The source port identifies traffic originating from the server.
- **Destination port number** The destination port identifies traffic originating from the client to the server.

Type of service and priority or precedence The IPv4 header has 3-bit precedence field so there are 8 possible levels [42]. In addition there is type of service field which originally was a bit pattern [42], then enumeration value [2] and now together with precedence filed a 8-bit Differentiated Services Field (DS Field) [34].

In IPv6 there are was originally 4-bit priority field (8 levels for real time traffic and 8 levels for elastic traffic) [17] but it is now replaced with 8-bit traffic class field (DS Field) [18, 34].

Source host address We can identify end system sending data and based on that classify traffic (we can identify the customer).

Destination host address We can identify end system receiving data.

Flow A flow is defined as a sequence of packets which have some common denominator. Depending on granularity it can be anything between:

- **Source and destination network** Packets between two networks share same routing information (in single-class routing). In practise, the addresses have the same prefix (source and destination individually).
- **5-tuple** The most fine-grained descriptor for a flow is 5-tuple, which consists of source and destination addresses, transport protocol identifier and source and destination ports, for example (130.233.154.145, 130.233.228.32, 6 (TCP), 33877, 80).

It is also possible, that some other descriptor is used to identify flow. In IPv6 there is defined a 20-bit flow label that can be used to identify flow with source host address [18].

[1] Of course, a protocol analysis could be done on traffic but in most cases that is not feasible.

3 Internet protocol

The Internet Protocol (IP) development started in DARPA ARPANET in 1969 using IMP nodes. As the network grow, the development work for common fault tolerant protocol started in 1974. The architecture and the core protocols was ready by end of 1970s and beginning of 1980s [40, 42, 41, 43]. In January 1983 the ARPANET changed from Network Control Protocol (NCP) to Internet Protocol family (TCP/IP). The most commonly used applications programming interface (API) for TCP/IP services was developed for BSD 4.2 UNIX in 1983: the same interface is used on most platforms [33].

The IP works roughly at network layer in ISO Open Systems Interconnection (OSI) model [27]. The supporting protocols are Internet Control Message Protocol (ICMP) [41] (error reporting, configuration and diagnosis), Internet Group Management Protocol (IGMP) [16, 20] (multicast management), and Address Resolution Protocol (ARP) [39] which performs mapping between link layer addresses and IP addresses if needed.

The IP is a datagram delivery service: each datagram contains enough information to carry it to its destination. There is no call setup: the service is *connectionless*. The network, however, does not make any guarantees if the datagram is delivered to destination. If the packet is lost, corrupted, misdelivered, or for some other reason not delivered to its intended destination, the network does nothing to recover from the failure. This service model is commonly called as *best effort* or *unreliable* service. IP does not either guarantee anything about order in which packets arrive to destination nor that packets are delivered at most once (datagrams may duplicate). However, there is limited life time for each datagram.

The best effort connectionless service is the simplest service a internetwork can provide: this makes it possible to transfer datagrams over any link layer technology (as in humorous "specification" in [49]). If the link losses some packets, that's fine. If the link delivers all packets, even better.

3.1 Transmission Control Protocol (TCP)

If we want to transport data reliably on top of IP, we need a transport protocol which hides unreliability of IP from application. The most simple solution is that the sender sends one data segment; if the receiver receives it successfully, it acknowledges received packet. As sender receives acknowledgment it may send the second segment. If either segment or acknowledgment is lost or receiver is not online, the sender does not receive acknowledgment in specified time and retransmits segment. [38]

This is not very efficient as the transmission speed is restricted by the round trip delay. A better approach is to use sliding window scheme: the receiver announces how much it is willing to receive: the sender can send this much without getting acknowledgment.

The Transmission Control Protocol (TCP) provides reliable byte stream using sliding window flow control scheme. The original scheme, however, did work badly in case of congestion as was seen first in 1986 in "congestion collapses" [30]. The scheme was then improved by introducing following methods:

Round-trip-time variance estimation Better estimates to find out when a segment is lost or when it is just late. In case of rising congestion event the delay will increase very much. The retransmit timer is set to value mean plus four times variation.

Exponential retransmit timer backoff Limit data rate sent to network to help clear out the congestion.

Slow-start Probe for available bandwidth.

More aggressive receiver ack policy Receiver acknowledges data as soon as possible to avoid retransmits.

Dynamic window sizing on congestion Adapts for changed situation on network.

Karn's clamped retransmit backoff Limit data rate.

Fast retransmit Fast recovery if only one segment is lost.

In addition to window used for flow control (e.g.. the sender does not overrun the receiver) the concept of *congestion window* was introduced. The congestion window tells sender how much it can send to network and the the sender selects minimum of those two windows.

The improvements made lead to more graceful operation in case of congestion. The problem is that there are difficulties to estimate (specially in case of retransmissions) round trip delay which is vital for

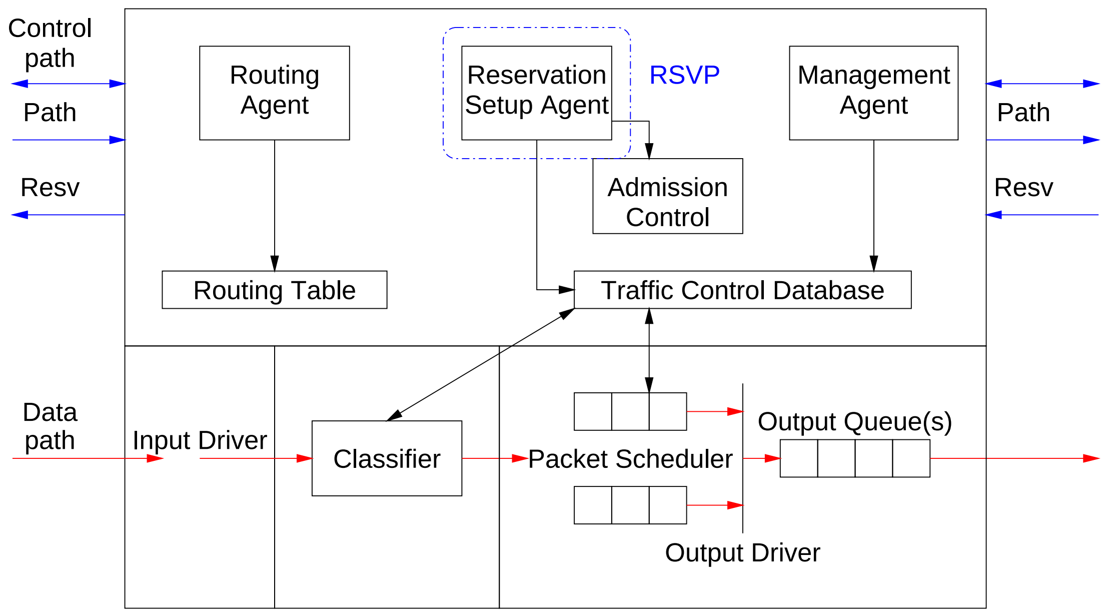

Figure 3: Integrated Services reference model with RSVP [13, 14]

maintaining steady flow. This has been addressed with timestamp option [12]. Current TCP congestion control algorithms and discussion can be found from [1, 22].

The TCP is *elastic* transport protocol: it adapts it transfer rate to available bandwidth on on network. It does not make any efforts to have minimum rate but only deliveries data in reliable manner.

3.2 User Datagram Service (UDP)

The User Datagram Service (UDP) provides program addressing (ports) and optional data integrity check (checksum for payload) [40]. It does not add any delivery-reliability, hence the original name "Unreliable".

The UDP is suitable for use as a carrier for real time traffic as it does not have any flow control or retransmissions which could affect on timing. With real-time traffic retransmissions are generally useless as retransmitted data arrives too late to be for any use.

3.2.1 Real-time Transport Protocol

For a real-time application, there is more need for control than the UDP provides. The real-time transport protocol (RTP) and accompanying real-time control protocol (RTCP) are designed for this purpose [46]. The RTP packets encapsulated in UDP packets carry the actual real-time data and have a sequence number and a time stamp. The sequence number makes possible that the receiver can find out if there are some dropped packets. The timestamp is used to detect jitter introduced by network and end systems.

4 Integrated Services

The lack of any QoS guarantees or levels in Internet is considered as one of main limitation of more wide use of Internet. To solve this problem an IETF Internet Integrated Services working group was formed. It defined a framework for resource reservation and performance guarantees. This framework is independent from protocols used for signaling and implementation details [13].

Each node is divided into two parts: background process and traffic forwarding (see Figure 3). The background process takes care of routing, reservation setups and admission control in addition to management. The traffic forwarding part classifies traffic based on information on traffic control database and based on this classification it is scheduled to right queues.

By now there are two service classes. Both of them support merging of flows for scalability and have rules to substitute "as good or better" service.

4.1 Guaranteed Quality of Service

The guaranteed service is designed for applications which require certain minimum bandwidth and maximum delay. The service, as it provides firm (mathematically provable) bounds for end-to-end queuing delay, makes possible to provide service that guarantees both delay and bandwidth. [47]

The traffic is considered as fluid model: delivered queuing delays do not exceed the fluid delays by more than the specified error bounds. The maximum delay is

$$d_{\text{max}} = \frac{b-M}{R}\frac{p-R}{p-r} + \frac{M+C_{\text{tot}}}{R+D_{\text{tot}}}, p > R \geq r \tag{3}$$

$$d_{\text{max}} = \frac{M+C_{\text{tot}}}{R+D_{\text{tot}}}, r \leq p \leq R \tag{4}$$

where b is a token bucket depth, r is a bucket rate, p is a token bucket plus peak rate, M is a maximum datagram size, R is a bandwidth allocated to connection, C_{tot} is a end-to-end sum of rate-dependent error terms, and D_{tot} is a end-to-end sum of rate-independent, per-element error terms. When the resource reservation is being made, each node calculates its values for C and D.

As long as the traffic is conforming to traffic specification (TSpec: b, r, p, m (a minimum policed unit; used to estimate link level overhead), M) the network element must transmit the packets conforming to receiver specification (RSpec: R, S (a slack term; "extra" time the node can delay the datagram)). If the traffic exceeds the traffic specification, the non-conforming datagrams must be considered as best-effort datagrams. They should not be given any presence over other best-effort datagrams (to avoid misuse) nor they should be discarded as erroneous packets as the originally conforming traffic may become non-conforming in network.

4.2 Controlled-load Network Element Service

The controlled-load service provides independent the network element load the client data flow with QoS closely approximating the QoS the flow would receive in unloaded[2] network. It uses capacity (admission) control to assure this. [50]

As in the guaranteed service (Chapter 4.1) the service is provided for a flow conforming the same TSpec. The applications may assume that only very few if any packets are lost and only very few if any packets greatly exceed minimum transit delay. If a non-conforming packet is received, the network element must ensure that

1. the other controlled-load flows receive expected QoS,
2. the excess traffic does not unfairly impact on best-effort traffic,
3. the excess traffic is delivered best-effort basis if sufficient resources exists.

4.3 Resource Reservation Protocol (RSVP)

All resource reservation systems need a setup (signaling) protocol to allocate needed resources from the network. For the IP networks, the Resource ReSerVation Protocol (RSVP) is specified [14]. The RSVP is independent from Integrated Services model and can be used with variety of QoS services as the Integrated Services can be used with variety of setup protocols [51].

The RSVP is receiver initiated: this provides better scalability for large multicast receiver groups, more flexible group membership and diverse receiver requirements. The RSVP sender sends *Path* message which records the route packets travel to receiver and has traffic characterization information. On reception of Path message the receiver sends *Resv* message to reserve capacity from the network. This message travels hop-by-hop same route (other direction) the Path message traveled.

The RSVP supports three types of reservations:

Wildcard-Filter (WF) The WF reservation is shared with all senders: it is propagated towards all senders and is extended automaticly to new senders as they appear.

Fixed-Filter (FF) The reservation is distinct (not shared between senders) and the sender is specified explicitly.

[2] *Unloaded* as meaning *not heavily loaded or congested*, **not** *no other traffic*.

Table 3: Integrated Services and ATM QoS mapping [19]

Integrated Service Class	ATM Service Class
Guaranteed Service	CBR or rt-VBR
Controlled Load	nrt-VBR or ABR (with minimum rate MCR)
Best Effort	UBR or ABR

Shared-Explicit (SE) The reservation is shared by selected senders. Compared to WF reservation, the receiver can select set of senders.

One of crucial problems is resource reservation charging. If the resources can be reserved without charge, all network users will want reserve all of bandwidth to themselves. For network to provide QoS, there must be some monetary cost associated to reservation which corresponds to amount of resources reserved. There is no currently any billing mechanism defined, but there is mechanism for cryptographic authentication [7]. The key management and accounting [4] are still big issues to be solved.

4.4 QoS routing

In larger networks there are generally several alternative routes between two hosts. These alternative routes provide added fault tolerance and possibility to share network load. The routing algorithms are based on graph theory and practical implementations within one administrative domain are based either on distance vector or more commonly today to topology and link state, for example the Open Shortest-Path First (OSPF).

If we include QoS parameters, for example needed bandwidth or delay bounds, the picture changes. We must take those considerations into account when calculating network topology. The ATM Private Network-Network Interface (PNNI) [15] routing information uses source traffic descriptor to determine which links to include into topology and which not. There can be great number of different topologies depending on requirements so handling of those can be a difficult task.

Another factor where QoS constrains affect on routing are changes in route tables in life time of connection. If the network and routing algorithm are not stabile, the route may change very often, specially if the network is congested. If routing information is changed after reservation setup there is no resources reserved for this particular connection in new route. For this reason, the RSVP requires that route for certain session is pinned: the concept is called as *path pinning* and it is relied on periodic RSVP Path updates to change reservation to the new route [3].

4.5 Integrated Services and ATM

The IP and ATM world have one basic difference: the ATM is connection oriented as the IP is connectionless. There are two basic ways to realize IP-over-ATM: using permanent virtual circuits (PVC) which emulate point-to-point links (leased lines) and switched virtual circuits which are set up on-demand. The reservation in ATM is “hard state” (active as long as it is not released) while on RSVP the reservation is “soft state” (active as long there are periodic updates).

The mapping of Integrated services to ATM service classes is quite straightforward and is presented in Table 3. The ATM traffic descriptors (PCR, SCR, MBS) are set to values based on peak rate, bucket depth, RSpec, and Receiver TSpec.

4.6 Bandwidth allocation in subnets

For point-to-point links the bandwidth allocation and prioritizing is done by router. In multipoint networks where may be some shared media each host can send as much as they want. Some mechanism is needed to make sure that bandwidth allocations do not exceed available bandwidth in network. This is a task for bandwidth broker or Subnet Bandwidth Manager (SBM) [52].

Each (RVSP) request for bandwidth goes through SBM and if sufficient capacity in network exists, it grants request. It does not, however, policy requests in any way, it is just a book keeper.

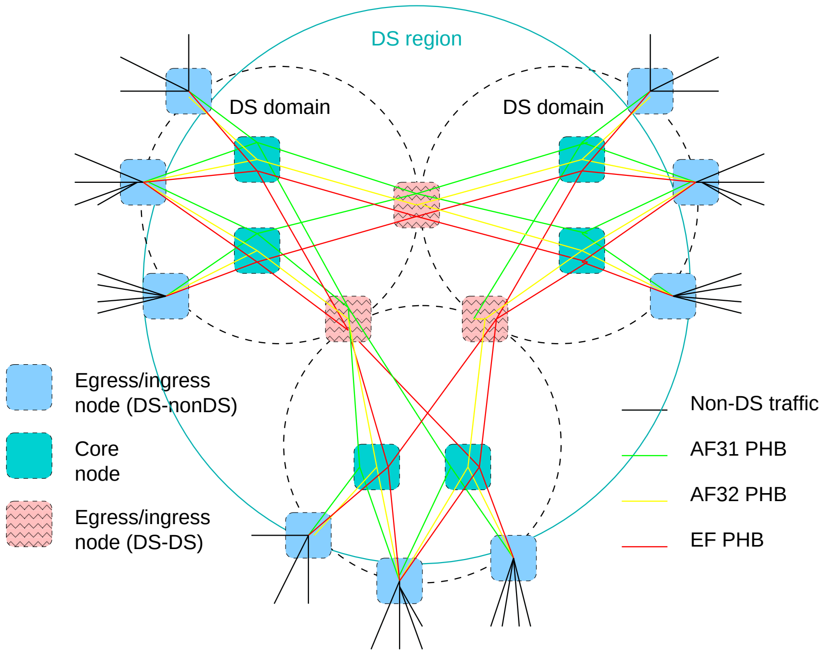

Figure 4: Differentiated service network [11, 10].

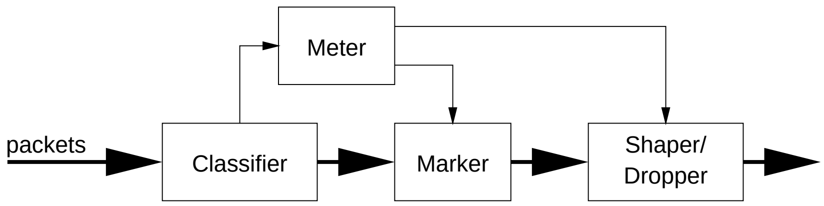

Figure 5: Logical view of a Packet Classifier and Traffic Conditioner in differentiated service edge node [11].

5 Differentiated Services

One of main problems with any resource reservation technology is burden needed for maintaining state information for each flow. As in come central points the number of simultaneous flows may exceed hundred thousand. If we estimate that each flow lasts for 10 seconds, there comes and goes more than 10,000 flows per second. For a reference, a large telephone exchange can handle up to million BHCA (busy hour call attempts) which equals to few hundred calls per second on average.

The number of flows makes maintaining per flow state information infeasible in core routers. Time needed to look up database entry for 5-tuple in each packet is considerable overhead compared for normal destination address lookup from routing table. The solution is to use per-packet stateless information.

For differentiated service approach several requirements were identified and addressed [11]. The key requirements are:

- Independence of applications, services and policing.
- Deployable incremental (only some part(s) of path), interoperability with QoS other technologies,
- No customer or microflow information or state in core network nodes – no nop-by-hop signaling. Core nodes utilize only small set of simple aggregated classification policies.

The differentiated service (DS) architecture is presented in Figure 4. The traffic is classified at edges of network (Figure 5). Each datagram is possibly conditioned and assigned to one of behavior aggregates which is identified by DS codepoint. At the core of the network, packets are forwarded according to the per-hop behavior (PHB) associated with the DS codepoint.

A customer (possibly other network operator) makes service level agreement (SLA) with the network operator. The SLA can be either *qualitative* ("Traffic offered at service level A will be delivered with low latency") or *quantitative* ("90% of in profile traffic delivered at service level B will experience no more than 50 ms latency") [10].

Based on SLAs the network provider assigns proper service level specifications (SLS) to boundary nodes. The nodes have four components as seen in Figure 5. The meters measure if submitted traffic conforms to a profile. Based on measurements other components will implement the policing. Markers re-mark traffic: to demote out-of-profile traffic to different PHB, to conform SLS codepoint mutation, and to ensure that only valid codepoints are used. Shapers delay traffic so that it does not exceed profile and droppers discard non-conforming traffic. [10]

5.1 Per-Hop-Behavior Groups

The per-hop-behavior groups are the actual mechanism to implement needed service differentiation in core networks. There should not be too many PHB groups as it complicates efficient router design [11]. Currently there are proposals for two PHB groups:

Assured Forwarding PHB Group (AF) provides four *independently* forwarded traffic classes, each with three drop precedences. A single DS node does not reorder packets of the same microflow if they belong into the same AF class. Note that this does not guarantee that packets do not get reordered as they travel through the network as datagrams may take different path as routing information changes. [24]

Each class is assigned some partition of bandwidth and buffer capacity. One way to use classes is "Olympic service": packets are assigned to "gold", "silver", and "bronze" classes. The "gold" class has lighter load than the other two classes. The customer may select one of these classes (which each are of of different cost).

An Expedited Forwarding PHB Group (EF) can be used to build a low loss, low latency, low jitter assured bandwidth, end-to-end service through DS domain [31]. This makes possible to provide end-to-end "virtual leased lines" or Premium service [35].

5.2 Service Examples

In [10] there are defined some service examples. Codepoints used in examples are not officially assigned.

Better than Best-Effort (BBE) provides service prioritized to best-effort. The traffic conforming contract (for example: 1 Mbps, any egress point) is marked with AF11 mark and traffic non-conforming is marked with AF13. For example the web server provider can use this to provide better performance to its clients.

Leased Line Emulation uses EF to implement this. A traffic contract "1 Mbps, egress point B, discard non-conforming" provides virtual 1 Mbps leased line to destination at egress point B.

Quantitative Assured Media Playback is similar to Leased Line Emulation but bursts are allowed and no traffic is (imminently) discarded. Traffic not exceeding basic rate (like 100 kbps) is marked with AF11, burst at maximum rate 200 kbps and size up to 100 kB are marked with AF12 and larger bursts with AF13.

5.3 Use of RSVP with Differentiated Services

It is commonly agreed that RSVP and integrated services do not provide enough scalability in high speed core networks. The differentiated services on the other hand may not have enough granularity to work just on few flows resulting non-guaranteed service in access network. One solution provided for that is using RSVP/intserv at edges of Internet (access networks) and use differentiated services on core networks [9].

The differentiated service networks are seen by RSVP/intserv connections as a single hop. The RSVP/intserv network edge nodes and diffserv network border nodes take care of mapping RSVP requests and flows to proper differentiated services PHB group.

6 Special Considerations

6.1 IP Security

The use of the IP Security protocols [6] causes some problems for both for the RSVP and differentiated service.

If Encapsulating Security Payload (ESP) [5] is used, the upper protocol layers are encrypted (discussion about tunneling mode see Section 6.2) and the network nodes cannot know the port numbers or protocols. The RSVP uses port numbers to make difference between different flows between two hosts (for example one flow for data comminíсatXXX and one for audio transfer).

As the transport layer is encrypted the network nodes cannot know the ports and thus cannot differentiate between flows requiring real time handling and bulk transfers. If the datagram is only authenticated the port numbers are visible but they are on different position which may cause performance problems on router.

This problem was solved by introducing “virtual destination port which is actually the IP SEC Security Parameter Index (SPI). We must then make sure that flows need different QoS have different SPI [8].

If services are classified for differentiated services networks based on port numbers [26], the encryption hides needed information. In this case the end system should be able to tell network what kind of service each packet needs.

6.2 Tunneling

There are three main uses for use tunneling: the first is to build (possibly secure) virtual private networks (VPN), the second one is to provide transport for protocols the network between does not support (as currently IPv6). The third use is tunneling subscriber traffic from access server to Internet service provider (ISP). The access server can be located in local telephone exchanges so Internet connections do not reserve circuit switched capacity from telephone network.

As the original IP packets are encapsulated to IP packets, the port numbers or DS code points are not visible to routing nodes. This problem has been solved for RSVP by using IP-in-UDP encapsulation. UDP source ports are used to identify individual (or aggregated) flows. The RSVP reservations are tunneled and corresponding reservations are made for tunnel also. [48]

For differentiated services the solution is simpler: the DS codepoint is copied to IP datagram which carries the original IP datagram. This way the datagram will receive same service as the original would. Packet reordering may cause problems with some tunneling protocols so packets in same flow should have same DS code.

7 Conclusions

The well known best effort service is not satisfactory for real time applications. The Internet pricing is currently in many cases flat rate: a monthly fee or a fee based on time on-line. The user has no possibility to get better service even if he is willing to pay more for premium service.

Currently there are two main efforts to provide control of QoS: the Integrated Services and the Differentiated Services. The first one is based on resource reservation: the main advantage is possibility to get well defined QoS, the main disadvantage is the need to maintain connection state in all network nodes between end systems. It also needs support from both end systems and from all networks between these two systems to be useful at all.

The Differentiated Services model is based on using per-hop behaviors which are marked as DS code points to IP datagram. The most important advantage is that the decision at core network node is local both in space and in time which guarantees scalability. It can also deployed step by step and interoperability is maintained. The disadvantage is that no connection admission control is done which may cause temporary overload situations on low bandwidth connections.

It looks like there are two possible ways to implement QoS: a combination where diffserv is used in core networks and RSVP/intserv in access network or diffserv-only network. If the applications do not soon support RSVP, the latter alternative is more probable.

References

[1] M. Allman, V. Paxon, and W. Stevens. TCP congestion control. Request for Comments (Standards Track) RFC 2581, Internet Engineering Task Force, April 1999. (Obsoletes RFC 2001). URL:http://www.ietf.org/rfc/rfc2581.txt.

[2] P. Almquist. Type of service in the internet protocol suite. Request for Comments (Standards Track) RFC 1349, Internet Engineering Task Force, July 1992. URL:http://www.ietf.org/rfc/rfc1349.txt.

[3] G. Apostolopoulos, R. Guerin, S. Kamat, A. Orda, T. Przygienda, and D. Williams. QoS routing mechanisms and OSPF extensions. Internet Draft draft-guerin-qos-routing-ospf-05, Internet Engineering Task Force, April 1998. Expires October 1999. Work in Progress. URL:http://www.ietf.org/internet-drafts/draft-guerin-qos-routing-ospf-05.txt.

[4] Jari Arkko. Requirements for internet-scale accounting management. Internet Draft draft-arkko-acctreq-00, Internet Engineering Task Force, August 1998. Work in Progress. URL:http://www.ietf.org/internet-drafts/draft-ietf-rsvp-md5-08.txt.

[5] R. Atkinson. IP encapsulating security payload (ESP). Request for Comments (Standards Track) RFC 1827, Internet Engineering Task Force, August 1995. URL:http://www.ietf.org/rfc/rfc1827.txt.

[6] R. Atkinson. Security architecture for the internet protocol. Request for Comments (Standards Track) RFC 1825, Internet Engineering Task Force, August 1995. URL:http://www.ietf.org/rfc/rfc1825.txt.

[7] Fred Baker, Bob Lindel, and Mohit Talwar. RSVP cryptographic authentication. Internet Draft draft-ietf-rsvp-md5-08, Internet Engineering Task Force, February 1999. Expires August 1999. Work in Progress. URL:http://www.ietf.org/internet-drafts/draft-ietf-rsvp-md5-08.txt.

[8] L. Berger and T. O'Malley. (rsvp) extensions for IPSEC data flows. Request for Comments (Standards Track) RFC 2207, Internet Engineering Task Force, September 1997. URL:http://www.ietf.org/rfc/rfc2207.txt.

[9] Y. Bernet, R. Yavatkar, P. Ford, F. Baker, L. Zhang, M. Speer, and R. Braden. Interoperation of RSVP/intserv and diffserv networks. Internet Draft draft-ietf-issll-diffserv-rsvp-01, Internet Engineering Task Force, March 1999. Expires September, 1999. Work in progress. URL:http://www.ietf.org/in-notes/draft-ietf-issll-diffserv-rsvp-01.

[10] Yoram Bernet, James Binder, Steven Blake, Mark Carlson, Brian E. Carpenter, Srinivasan Keshav, Elwyn Davies, Borje Ohlman, Dinesh Verma, Zheng Wang, and Walter Weiss. A framework for differentiated services. Internet Draft draft-ietf-diffserv-framework-02, Internet Engineering Task Force, February 1999. Expires September 1999. Work in Progress. URL:http://www.ietf.org/internet-drafts/draft-ietf-diffserv-framework-02.txt.

[11] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss. An architecture for differentiated services. Request for Comments (Infromational) RFC 2475, Internet Engineering Task Force, December 1998. URL:http://www.ietf.org/rfc/rfc2475.txt.

[12] D. Borman, R. Braden, and V. Jacobson. TCP extensions for high performance. Request for Comments (Standards Track) RFC 1323, Internet Engineering Task Force, May 1992. (Obsoletes RFC 1185). URL:http://www.ietf.org/rfc/rfc1323.txt.

[13] R. Braden, D. Clark, and S. Shenker. Integrated services in the internet architecture: an overview. Request for Comments (Informational) RFC 1633, Internet Engineering Task Force, June 1994. URL:http://www.ietf.org/rfc/rfc1633.ps.

[14] R. Braden (Ed), L. Zhang, S. Berson, S. Herzog, and S. Jamin. Resource reservation protocol (rsvp) – version 1 functional specification. Request for Comments (Standards Track) RFC 2205, Internet Engineering Task Force, September 1997. URL:http://www.ietf.org/rfc/rfc2205.txt.

[15] Technical Committee. Private network-network interface specification version 1.0 (PNNI 1.0). Technical Report af-pnni-0055.000, The ATM Forum, March 1996. URL:ftp://ftp.atmforum.com/pub/approved-specs/af-pnni-0055.000.pdf.

[16] S. Deering. Host extensions for IP multicasting. Request for Comments (Standard) STD 5, RFC 1112, Internet Engineering Task Force, August 1989. (Obsoletes RFC 988). URL:http://www.ietf.org/rfc/rfc1112.txt.

[17] S. Deering and R. Hinden. Internet protocol, version 6 (IPv6) specification. Request for Comments (Standards Track) RFC 1883, Internet Engineering Task Force, January 1996. (Obsoleted by RFC 2460). URL:http://www.ietf.org/rfc/rfc1883.txt.

[18] S. Deering and R. Hinden. Internet protocol, version 6 (IPv6) specification. Request for Comments (Standards Track) RFC 2460, Internet Engineering Task Force, December 1998. (Obsoletes RFC 1883). URL:http://www.ietf.org/rfc/rfc2460.txt.

[19] E. Crawley (Ed), L. Berger, S. Berson, F. Baker, M. Borden, and J. Krawczyk. A framework for integrated services and RSVP over ATM. Request for Comments (Informational) RFC 2382, Internet Engineering Task Force, August 1998. URL:http://www.ietf.org/rfc/rfc2382.txt.

[20] W. Fenner. Internet group management protocol, version 2. Request for Comments (Standards Track) RFC 2236, Internet Engineering Task Force, November 1997. (Updates RFC 1112). URL:http://www.ietf.org/rfc/rfc2236.txt.

[21] Paul Ferguson and Geoff Huston. *Quality of Service: delivering QoS on the Internet and in corporate networks*. John Wiley & Sons, 1998.

[22] S. Floyd and T. Henderson. The NewReno modification to TCP's fast recovery algorithm. Request for Comments (Experimental) RFC 2582, Internet Engineering Task Force, April 1999. URL:http://www.ietf.org/rfc/rfc2581.txt.

[23] G. Gosztony, K. Rahko, and R. Chapuis. The grade of service in the world wide telephone network. *Telecommunication Journal*, 46(IX):556 – 565, 1979. The second part published in October issue.

[24] Juha Heinanen, Fred Baker, Walter Weiss, and John Wroclawski. Assured forwarding PHB group. Internet Draft draft-ietf-diffserv-af-06, Internet Engineering Task Force, February 1999. Expires October 1999. Work in Progress. URL:http://www.ietf.org/internet-drafts/draft-ietf-diffserv-af-06.txt.

[25] Internet protocol data communication service - IP packet transfer and availability performance parameters. Draft New ITU-T Recommendation I.380, International Telecommunication Union, June 1998. Work in progress.

[26] Mika Ilvesmäki and Marko Luoma. Multiclass differentiation of internet traffic using learning vector quantization. Submitted for publication, 1999.

[27] ISO. ISO/IEC 7498-1:1994 information technology – open systems interconnection – basic reference model: The basic model. ISO standard ISO/IEC 7498-1, ISO / JTC 1 / SC 21, 1994.

[28] ITU-T. Grades of service in digital international telephone exchenges. ITU-T Recommendation E.543, International Telecommunication Union, 1988.

[29] ITU-T. Network grade of service parameters and target values for circuit-switched services in the evolving ISDN. ITU-T Recommendation E.721, International Telecommunication Union, 1991.

[30] V. Jacobson. Congestion avoidance and control. In *Proceedings of the ACM SIGCOMM Conference*, pages 314–329, August 1988.

[31] Van Jacobson, Kathleen Nichols, and Kedarnath Poduri. An expedited forwarding PHB. Internet Draft draft-ietf-diffserv-phb-ef-02, Internet Engineering Task Force, February 1999. Expires August 1999. Work in Progress. URL:http://www.ietf.org/internet-drafts/draft-ietf-diffserv-phb-ef-02.txt.

[32] Marko Luoma, Markus Peuhkuri, and Tomi Yletyinen. Quality of service for IP voice services - is it necessary? In *Proceedings of Voice, Video and Data '98*, pages 242–253. SPIE, November 1998.

[33] Marshall Kirk McKusick, Keith Bostic, Michael J. Karels, and John S. Quarterman. *The Design and Implementation of the 4.4BSD Operating System.* Addison-Wesley Publishing Company, Inc., 1996.

[34] K. Nichols, S. Blake, F. Baker, and D. Black. Definition of the differentiated services field (DS field) in the IPv4 and IPv6 headers. Request for Comments (Standards Track) RFC 2474, Internet Engineering Task Force, December 1998. (Obsoletes RFC 1455 and RFC 1349). URL:http://www.ietf.org/rfc/rfc2474.txt.

[35] K. Nichols, V. Jacobson, and L. Zhang. A two-bit differentiated service architecture for the internet. Internet Draft draft-nichols-diff-svc-arch-00, Internet Engineering Task Force, November 1997. Work in Progress. URL:ftp://ftp.ee.lbl.gov/papers/dsarch.pdf.

[36] Vern Paxon and Sally Floyd. Why we don't know how to simulate the internet. In *Proceedings of the 1997 Winter Simulation Conference*, December 1997.

[37] V. Paxson, G. Almes, J. Mahdavi, and M. Mathis. Framework for IP performance metrics. Request for Comments (Informational) RFC 2330, Internet Engineering Task Force, May 1998. URL:http://www.ietf.org/rfc/rfc2330.txt.

[38] Larry L. Peterson and Bruce S. Davie. *Computer Networks: A Systems Approach.* Morgan Kaufman Publishers, Inc., 1996.

[39] D. Plummer. Ethernet address resolution protocol: Or converting network protocol addresses to 48.bit ethernet address for transmission on ethernet hardware. Request for Comments (Standard) RFC 826, Internet Engineering Task Force, November 1982. URL:http://www.ietf.org/rfc/rfc826.txt.

[40] J. Postel. User datagram protocol. Request for Comments (Standard) STD 6, RFC 768, Internet Engineering Task Force, August 1980. URL:http://www.ietf.org/rfc/rfc768.txt.

[41] J. Postel. Internet control message protocol. Request for Comments (Standard) STD 5, RFC 792, Internet Engineering Task Force, September 1981. (Obsoletes RFC 777). URL:http://www.ietf.org/rfc/rfc792.txt.

[42] J. Postel. Internet protocol. Request for Comments (Standard) RFC 791, Internet Engineering Task Force, September 1981. (Obsoletes RFC 760). URL:http://www.ietf.org/rfc/rfc791.txt.

[43] J. Postel. Transmission control protocol. Request for Comments (Standard) STD 7, RFC 793, Internet Engineering Task Force, September 1981. URL:http://www.ietf.org/rfc/rfc793.txt.

[44] J. Postel and J. Reynolds. Telnet protocol specification. Request for Comments (Standard) STD 8, RFC 854, Internet Engineering Task Force, May 1983. (Obsoletes RFC 764). URL:http://www.ietf.org/rfc/rfc854.txt.

[45] Kauko Rahko, Stefan Herzberg, and Timo Rahko. Grade of service and human factors. In *6th International Symposium on Human Factors in Telecommunications*, Stockholm, 1972.

[46] H. Schulzrinne, S. Casner, R. Frederik, and V. Jacobson. RTP: A transport protocol for real-time applications. Request for Comments (Standards Track) RFC 1889, Internet Engineering Task Force, January 1996. URL:http://www.ietf.org/rfc/rfc1889.txt.

[47] S. Shenker, C. Partridge, and Guerin R. Specification of guaranteed quality of service. Request for Comments (Standards Track) RFC 2212, Internet Engineering Task Force, September 1997.

[48] A. Terzis, J. Krawczyk, J. Wroclawski, and L. Zhang. RSVP operation over ip tunnels. Internet Draft draft-ietf-rsvp-tunnel-03, Internet Engineering Task Force, April 1999. Expires October, 1999. Work in progress. URL:http://www.ietf.org/in-notes/draft-ietf-rsvp-tunnel-03.txt.

[49] D. Waitzman. A standard for the transmission of IP datagrams on avian carriers. RFC 1149, Internet Engineering Task Force, April 1990. URL:http://www.ietf.org/rfc/rfc1149.txt.

[50] J. Wroclawski. Specification of the controlled-load network element service. Request for Comments (Standards Track) RFC 2211, Internet Engineering Task Force, September 1997. URL:http://www.ietf.org/rfc/rfc2211.txt.

[51] J. Wroclawski. The use of RSVP with IETF integrated services. Request for Comments (Standards Track) RFC 2210, Internet Engineering Task Force, September 1997. URL:http://www.ietf.org/rfc/rfc2210.txt.

[52] Raj Yavatkar, Don Hoffman, Yoram Bernet, Fred Baker, and Michael Speer. SBM (subnet bandwidth manager): A protocol for RSVP-based admission control over IEEE 802-style networks. Internet Draft draft-ietf-issll-is802-sbm-07, Internet Engineering Task Force, November 1998. Work in Progress. URL:http://www.ietf.org/internet-drafts/draft-ietf-issll-is802-sbm-07.txt.